# Supplementary Information for Learning Risk Preferences Through Social Interaction: An Active Inference Approach

**Amir Hossein Tehrani-Safa**[1,2], **Reza Ghaderi**[1], **and Atiye Sarabi-Jamab**[3,*]

[1]Institute for Cognitive and Brain Sciences, Shahid Beheshti University, Tehran, P.O. Box 19839-63113, Iran.
[2]School of Cognitive Sciences, Institute for Research in Fundamental Sciences (IPM),Tehran, P.O. Box 19395-5746, Iran
[3]Faculty of Governance, University of Tehran, Tehran, 1417633461, Iran
[*]asarabi@ut.ac.ir

## ABSTRACT

This supplementary file provides additional details for the manuscript titled "Learning Risk Preferences Through Social Interaction: An Active Inference Approach". It includes three appendices containing mathematical derivations for equations presented in the main text, as well as a supplementary figure.

## Appendix S1: Simplifying the Variational energy

It is possible to simplify the expression $\ln\left[1-S\left(F(\rho^{(k)}\right)\right]$ in the following manner:

$$\begin{aligned}\ln\left[1-S\left(F(\rho^{(k)})\right)\right] &= \ln\left[1-\frac{1}{1+exp\left(-\beta\times F(\rho^{(k)})\right)}\right]\\ &= \ln\left[\frac{\exp\left(-\beta\times F(\rho^{(k)})\right)}{1+\exp\left(-\beta\times F(\rho^{(k)})\right)}\right]\\ &= \ln\left[\exp\left(-\beta\times F(\rho^{(k)})\right)\right]+\ln\left[\frac{1}{1+\exp\left(-\beta\times F(\rho^{(k)})\right)}\right]\\ &= -\beta\times F(\rho^{(k)})+\ln\left[S\left(F(\rho^{(k)})\right)\right]\end{aligned} \tag{S1.1}$$

Substituting the value of $\ln[1-S\left(F(\rho^{(k)})\right)]$ in Variational energy formula and carrying out the necessary operations yields the following expression:

$$\begin{aligned}I(\rho^{(k)}) &= C^{(k)}\times\ln\left[S\left(F(\rho^{(k)})\right)\right]+(1-C^{(k)})\times\left(-\beta\times F(\rho^{(k)})+\ln\left[S\left(F(\rho^{(k)})\right)\right]\right)\\ &\quad-\frac{\left(\rho^{(k)}-\mu_\rho^{(k-1)}\right)^2}{2\sigma_\rho^{(k-1)}}\\ &= \ln\left[S\left(F(\rho^{(k)})\right)\right]+\beta\times(C^{(k)}-1)\times F(\rho^{(k)})-\frac{\left(\rho^{(k)}-\mu_\rho^{(k-1)}\right)^2}{2\sigma_\rho^{(k-1)}}\end{aligned} \tag{S1.2}$$

## Appendix S2: Obtaining a quadratic approximation of the Variational energy $I(\rho^{(k)})$

Here one can observe the second order approximation of the Variational energy, resulting in an update equation. The equation obtained for $I(\rho^{(k)})$ contains non-quadratic terms:

$$I(\rho^{(k)}) = \ln\left[S\left(F(\rho^{(k)})\right)\right] + \beta \times (C_0^{(k)} - 1) \times F(\rho^{(k)}) - \frac{\left(\rho^{(k)} - \mu_\rho^{(k-1)}\right)^2}{2\sigma_\rho^{(k-1)}}$$

In order to approximate a Gaussian distribution, it is necessary to find a quadratic polynomial $\tilde{I}(\rho^{(k)})$ that approximates $I(\rho^{(k)})$. By doing so, the Gaussian distribution can be accurately represented.

$$\tilde{I}(\rho^{(k)}) = -\frac{\left(\rho^{(k)} - \mu_\rho^{(k)}\right)^2}{2\sigma_\rho^{(k)}} \tag{S2.1}$$

When the second order Taylor series of function $I(\rho^{(k)})$ is expanded around a particular expansion point, an approximate quadratic function $\tilde{I}(\rho^{(k)})$ can be obtained. This technique is referred to as Laplace approximation when the Taylor expansion is carried out at the peak of $I(\rho^{(k)})$[1]. An alternative site for the expansion point would be the mean value of $\rho^{(k)}$ that has been previously updated, symbolized by $\mu_\rho^{(k-1)}$[2]. The second suggestion is what we are going to use here.

Considering $\tilde{I}(\rho^{(k)})$ to be the second order Taylor expansion of $I(\rho^{(k)})$ about $\mu_\rho^{(k-1)}$, the first and second derivatives of both $\tilde{I}(\rho^{(k)})$ and $I(\rho^{(k)})$ are equivalent at this point:

$$\begin{aligned}\left(\frac{\partial I}{\partial \rho^{(k)}}\right)\Bigg|_{\rho^{(k)}=\mu_\rho^{(k-1)}} &= \left(\frac{\partial \tilde{I}}{\partial \rho^{(k)}}\right)\Bigg|_{\rho^{(k)}=\mu_\rho^{(k-1)}} \\ \left(\frac{\partial^2 I}{\partial (\rho^{(k)})^2}\right)\Bigg|_{\rho^{(k)}=\mu_\rho^{(k-1)}} &= \left(\frac{\partial^2 \tilde{I}}{\partial (\rho^{(k)})^2}\right)\Bigg|_{\rho^{(k)}=\mu_\rho^{(k-1)}}\end{aligned} \tag{S2.2}$$

The derivatives of $\tilde{I}(\rho^{(k)})$ at the point $\mu_\rho^{(k-1)}$ can be determined from Eq (S2.1). Substituting these values into Eq(S2.2) leads to the following results:

$$\begin{aligned}\left(\frac{\partial I}{\partial \rho^{(k)}}\right)\Bigg|_{\rho^{(k)}=\mu_\rho^{(k-1)}} &= \frac{\mu_\rho^{(k)} - \mu_\rho^{(k-1)}}{\sigma_\rho^{(k)}} \\ \left(\frac{\partial^2 I}{\partial (\rho^{(k)})^2}\right)\Bigg|_{\rho^{(k)}=\mu_\rho^{(k-1)}} &= -\frac{1}{\sigma_\rho^{(k)}}\end{aligned} \tag{S2.3}$$

Rearranging Eq(S2.3) yields into an updating equation for $\mu_\rho^{(k)}$ and $\sigma_\rho^{(k)}$:

$$\begin{aligned}\mu_\rho^{(k)} &= \mu_\rho^{(k-1)} + \sigma_\rho^{(k)} \times \left(\frac{\partial I}{\partial \rho^{(k)}}\right)\Bigg|_{\rho^{(k)}=\mu_\rho^{(k-1)}} \\ \frac{1}{\sigma_\rho^{(k)}} &= -\left(\frac{\partial^2 I}{\partial (\rho^{(k)})^2}\right)\Bigg|_{\rho^{(k)}=\mu_\rho^{(k-1)}}\end{aligned} \tag{S2.4}$$

The final step to obtain the explicit update equations for $\mu_\rho^{(k)}$ and $\sigma_\rho^{(k)}$ is to calculate the first and second derivatives of

$I(\rho^{(k)})$ with respect to $\mu_\rho^{(k-1)}$. We begin by computing the first derivative.

$$\begin{aligned}
\frac{\partial I}{\partial \rho^{(k)}} &= \frac{\partial}{\partial \rho^{(k)}} \left\{ \ln S\left(F(\rho^{(k)})\right) + \beta \times (C_o^{(k)} - 1) \times F(\rho^{(k)}) - \frac{\left(\rho^{(k)} - \mu_\rho^{(k-1)}\right)^2}{2\sigma_\rho^{(k-1)}} \right\} \\
&= \frac{\partial}{\partial \rho^{(k)}} \left\{ \ln S\left(F(\rho^{(k)})\right) + \beta \times (C_o^{(k)} - 1) \times F(\rho^{(k)}) \right\} - \frac{\partial}{\partial \rho^{(k)}} \left\{ \frac{\left(\rho^{(k)} - \mu_\rho^{(k-1)}\right)^2}{2\sigma_\rho^{(k-1)}} \right\} \\
&= \frac{\partial}{\partial \rho^{(k)}} \left\{ \ln S\left(F(\rho^{(k)})\right) \right\} + \beta \times (C_o^{(k)} - 1) \times \frac{\partial F(\rho^{(k)})}{\partial \rho^{(k)}} - \frac{\rho^{(k)} - \mu_\rho^{(k-1)}}{\sigma_\rho^{(k-1)}}
\end{aligned} \tag{S2.5}$$

Using a chain rule, it is possible to derive $\frac{\partial}{\partial \rho^{(k)}} \left\{ \ln \left[ S\left(F(\rho^{(k)})\right) \right] \right\}$ as follows:

$$\frac{\partial}{\partial \rho^{(k)}} \left\{ \ln \left[ S\left(F(\rho^{(k)})\right) \right] \right\} = \frac{1}{S\left(F(\rho^{(k)})\right)} \times \frac{\partial S}{\partial F(\rho^{(k)})} \times \frac{\partial F(\rho^{(k)})}{\partial \rho^{(k)}} \tag{S2.6}$$

where $\frac{\partial S}{\partial F(\rho^{(k)})}$ is:

$$\frac{\partial S}{\partial F(\rho^{(k)})} = \beta \times \left[ S\left(F(\rho^{(k)})\right) \right] \times \left[ 1 - S\left(F(\rho^{(k)})\right) \right] \tag{S2.7}$$

Finally we have:

$$\frac{\partial}{\partial \rho^{(k)}} \left\{ \ln \left[ S\left(F(\rho^{(k)})\right) \right] \right\} = \beta \times \frac{\partial F(\rho^{(k)})}{\partial \rho^{(k)}} \times \left[ 1 - S\left(F(\rho^{(k)})\right) \right] \tag{S2.8}$$

The following equation for $\frac{\partial I}{\partial \rho^{(k)}}$ can be calculated by substituting the value achieved for $\frac{\partial}{\partial \rho^{(k)}} \left\{ \ln \left[ S\left(F(\rho^{(k)})\right) \right] \right\}$ in the Eq (S2.5).

$$\begin{aligned}
\frac{\partial I}{\partial \rho^{(k)}} &= \beta \times \frac{\partial F(\rho^{(k)})}{\partial \rho^{(k)}} \left[ 1 - S\left(F(\rho^{(k)})\right) \right] + \beta \times \frac{\partial F(\rho^{(k)})}{\partial \rho^{(k)}} \times (C_o^k - 1) - \frac{\rho^{(k)} - \mu_\rho^{(k-1)}}{\sigma_\rho^{(k-1)}} \\
&= \beta \times \frac{\partial F(\rho^{(k)})}{\partial \rho^{(k)}} \times \left[ C_o^k - S\left(F(\rho^{(k)})\right) \right] - \frac{\rho^{(k)} - \mu_\rho^{(k-1)}}{\sigma_\rho^{(k-1)}}
\end{aligned} \tag{S2.9}$$

By differentiating Eq (S2.9), we can obtain the expression for $\frac{\partial^2 I}{\partial (\rho^{(k)})^2}$:

$$\begin{aligned}
\frac{\partial^2 I}{\partial (\rho^{(k)})^2} &= \frac{\partial}{\partial \rho^{(k)}} \left( \frac{\partial I}{\partial \rho^{(k)}} \right) \\
&= \frac{\partial}{\partial \rho^{(k)}} \left( \beta \times \frac{\partial F(\rho^{(k)})}{\partial \rho^{(k)}} \times \left[ C_o^k - S\left(F(\rho^{(k)})\right) \right] - \frac{\rho^{(k)} - \mu_\rho^{(k-1)}}{\sigma_\rho^{(k-1)}} \right) \\
&= \frac{\partial}{\partial \rho^{(k)}} \left( \beta \times \frac{\partial F(\rho^{(k)})}{\partial \rho^{(k)}} \times \left[ C_o^k - S\left(F(\rho^{(k)})\right) \right] \right) - \frac{\partial}{\partial \rho^{(k)}} \left( \frac{\rho^{(k)} - \mu_\rho^{(k-1)}}{\sigma_\rho^{(k-1)}} \right) \\
&= \beta \times \frac{\partial^2 F(\rho^{(k)})}{\partial (\rho^{(k)})^2} \times \left[ C_o^k - S\left(F(\rho^{(k)})\right) \right] - \beta \times \left( \frac{\partial F(\rho^{(k)})}{\partial \rho^{(k)}} \right)^2 \times \frac{\partial S}{\partial F(\rho^{(k)})} - \frac{1}{\sigma_\rho^{(k-1)}}
\end{aligned} \tag{S2.10}$$

Omitting the initial term which incorporates the second order derivative $\frac{\partial^2 F(\rho^{(k)})}{\partial (\rho^{(k)})^2}$ yields a more straightforward formula:

$$\frac{\partial^2 I}{\partial (\rho^{(k)})^2} \approxeq -\beta^2 \times \left( \frac{\partial F(\rho^{(k)})}{\partial \rho^{(k)}} \right)^2 \times \left[ S\left(F(\rho^{(k)})\right) \right] \times \left[ 1 - S\left(F(\rho^{(k)})\right) \right] - \frac{1}{\sigma_\rho^{(k-1)}} \tag{S2.11}$$

The value of $\frac{\partial^2 I}{\partial(\rho^{(k)})^2}$ is always kept at a negative level due to the presence of negative components. This ensures that $\sigma_\rho^{(k)}$ is maintained at a positive value, which is a necessary requirement based on its underlying characteristics.

Lastly, by substituting the values of $\frac{\partial I}{\partial \rho^{(k)}}$ and $\frac{\partial^2 I}{\partial(\rho^{(k)})^2}$ at the point $\rho^{(k)} = \mu_\rho^{(k-1)}$ into Equation Eq(S2.4), the update equation for $\mu_\rho^{(k)}$ and $\sigma_\rho^{(k)}$ can be obtained:

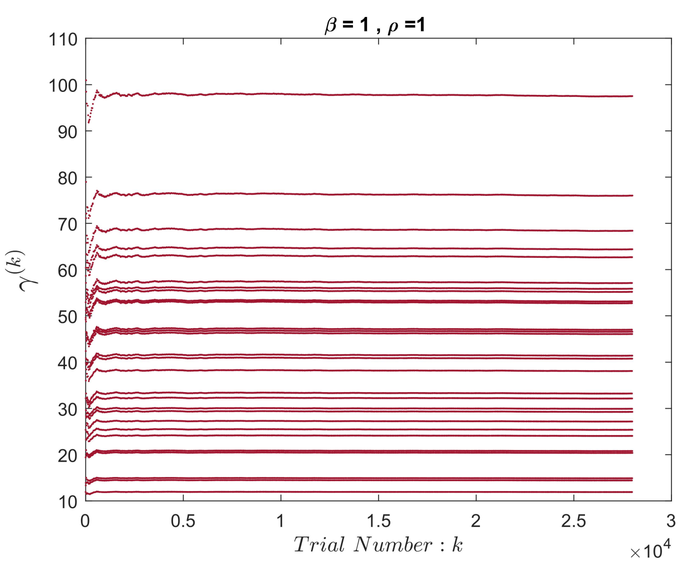
$$
\begin{aligned}
\mu_\rho^{(k)} &= \mu_\rho^{(k-1)} + \beta \times \left( \left. \frac{\partial F(\rho^{(k)})}{\partial \rho^{(k)}} \right|_{\rho^{(k)}=\mu_\rho^{(k-1)}} \right) \times \sigma_\rho^{(k)} \times \left( C_0^{(k)} - S\left(F(\mu_\rho^{(k-1)})\right)\right) \\
\frac{1}{\sigma_\rho^{(k)}} &= \frac{1}{\sigma_\rho^{(k-1)}} + \beta^2 \times \left( \left. \frac{\partial F(\rho^{(k)})}{\partial \rho^{(k)}} \right|_{\rho^{(k)}=\mu_\rho^{(k-1)}} \right)^2 \times \left[ S\left(F(\mu_\rho^{(k-1)})\right)\right] \times \left[1 - S\left(F(\mu_\rho^{(k-1)})\right)\right]
\end{aligned}
\tag{S2.12}
$$

Where the value of $\left. \frac{\partial F(\rho^{(k)})}{\partial \rho^{(k)}} \right|_{\rho^{(k)}=\mu_\rho^{(k-1)}}$ is:

$$
\begin{aligned}
\left. \frac{\partial F(\rho^{(k)})}{\partial \rho^{(k)}} \right|_{\rho^{(k)}=\mu_\rho^{(k-1)}} &= \frac{\partial}{\partial \rho^{(k)}} \left\{ p^k \times (r^k)^{\rho^k} - U_s \right\} \Bigg|_{\rho^{(k)}=\mu_\rho^{(k-1)}} \\
&= p^k \times \ln r^k \times (r^k)^{\mu_\rho^{(k-1)}}
\end{aligned}
\tag{S2.13}
$$

See method section in the main manuscript for the definition of function $F(\cdot)$.

## Appendix S3: Supplementary Figure 1




**Supplementary Figure 1. Variation of $\gamma^{(k)}$ across trials. As detailed in the main text, $\gamma^{(k)}$ is calculated as the derivative of $F(\rho^{(k)})$ with respect to $\rho^{(k)}$. The value of $\gamma^{(k)}$ is a function of the reward amount and the probability of the proposed gamble for each trial.**

## References


1. Friston, K., Mattout, J., Trujillo-Barreto, N., Ashburner, J. & Penny, W. Variational free energy and the laplace approximation. *Neuroimage* **34**, 220–234 (2007).
2. Mathys, C., Daunizeau, J., Friston, K. J. & Stephan, K. E. A bayesian foundation for individual learning under uncertainty. *Frontiers in human neuroscience* **5**, 39 (2011).